Non-Euclidean statistical analysis of covariance matrices and diffusion tensors

Dryden, Ian
University of South Carolina, Department of Statistics,
Columbia SC 29208, USA
E-mail: dryden@mailbox.sc.edu

Koloydenko, Alexey
Royal Holloway University of London, Department of Mathematics,
Egham, Surrey, TW20 0EX, UK
E-mail: a.koloydenko@rhul.ac.uk

Zhou, Diwei
University of Nottingham, School of Mathematical Sciences,
Nottingham, NG7 2RD, UK
E-mail: pmxdz@nottingham.ac.uk

Li, Bai
University of Nottingham, School of Computer Science,
Nottingham, NG8 1BB, UK
E-mail: bai@cs.nott.ac.uk

1 Introduction

The statistical analysis of covariance matrices occurs in many important applications, e.g. in diffusion tensor imaging and longitudinal data analysis. We consider the situation where it is of interest to estimate an average covariance matrix, describe its anisotropy, to carry out principal geodesic analysis and to interpolate between covariance matrices.

An important difference with standard statistical techniques is that non-Euclidean distances are most natural for comparing covariance matrices, which are symmetric, semi-positive definite matrices.

2 Diffusion tensors

In medical image analysis a particular type of covariance matrix arises in diffusion weighted imaging called a diffusion tensor. The diffusion tensor is a 3×3 covariance matrix which is estimated at each voxel in the brain, and is obtained by fitting a physically-motivated model on measurements from the Fourier transform of the molecule displacement density (Basser et al., 1994).

In the diffusion tensor model the water molecules at a voxel diffuse according to a multivariate normal model centred on the voxel and with covariance matrix Σ. The displacement of a water molecule $x \in \mathbf{R}^3$ has probability density function

$$f(x) = \frac{1}{(2\pi)^{3/2}|\Sigma|^{1/2}} \exp(-\frac{1}{2}x^T\Sigma^{-1}x).$$

The convention is to call $D = \Sigma/2$ the diffusion tensor, which is a symmetric positive semi-definite matrix. The diffusion tensor is estimated at each voxel in the image from the available MR images. The MR scanner has a set of magnetic field gradients applied at directions $g_1, g_2, \ldots, g_m \in RP^2$ with scanner gradient parameter b, where RP^2 is the real projective space of axial directions (with $g_j \equiv -g_j$,

$\|g_j\| = 1$). The data at a voxel consist of signals $(Z_0, Z_1, \ldots, Z_m)$ which are related to the Fourier transform of the displacement density in axial direction $g_j \in RP^2,\ j = 1, \ldots, m$, and the reading Z_0 is obtained with no gradient ($b = 0$). The Fourier transform in axial direction $g \in RP^2$ of the multivariate Gaussian displacement density is given by

$$\mathcal{F}(g) = \int \exp(i\sqrt{b} g^T x) f(x) dx = \exp(-b g^T D g),$$

and the theoretical model for the signals is

$$Z_j = Z_0 \mathcal{F}(g_j) = Z_0 \exp(-b g_j^T D g_j),\ \ j = 1, \ldots, m.$$

There is a variety of methods available for estimating D from the data $(Z_0, Z_1, \ldots, Z_m)$ at each voxel (see Alexander, 2005), including least squares regression and Bayesian estimation (e.g. Zhou et al., 2008). Noise models include log-Gaussian, Gaussian and more recently Rician noise (e.g. Fillard et al., 2007). A common method for visualizing a diffusion tensor is an ellipsoid with principal axes given by the eigenvectors of D, and lengths of axes proportional to $\sqrt{\lambda_i},\ i = 1, 2, 3$.

If a sample of diffusion tensors is available we may wish to estimate an average diffusion tensor matrix, investigate the structure of variability in diffusion tensors or interpolate at higher spatial resolution between two or more estimated diffusion tensor matrices.

A strongly anisotropic diffusion tensor indicates a strong direction of white matter fibre tracts, and plots of measures of anisotropy are very useful to neurologists. A measure that is very commonly used in diffusion tensor imaging is Fractional Anisotropy

$$FA = \left\{ \frac{k}{k-1} \sum_{i=1}^{k} (\lambda_i - \bar{\lambda})^2 / \sum_{i=1}^{k} \lambda_i^2 \right\}^{1/2}, \tag{1}$$

where $0 \leq FA \leq 1$ and λ_i are the eigenvalues of the diffusion tensor matrix. Note that $FA \approx 1$ if $\lambda_1 >> \lambda_i, i > 1$ (very strong principal axis) and $FA = 0$ for isotropy. In diffusion tensor imaging $k = 3$.

3 Non-Euclidean statistics

3.1 The Fréchet mean

When using a non-Euclidean distance $d()$ we must define what is meant by a 'mean covariance matrix'. Consider a probability distribution for a $k \times k$ covariance matrix S on a Riemannian metric space with density $f(S)$. The Fréchet (1948) mean Σ is defined as

$$\Sigma = \arg\inf_{\Sigma} \ \frac{1}{2} \int d(S, \Sigma)^2 f(S) dS,$$

and is also known as the Karcher mean (Karcher, 1977). The Fréchet mean need not be unique in general, although for many distributions it will be. Provided the distribution is supported only on the geodesic ball of radius r, such that the geodesic ball of radius $2r$ is regular (i.e. supremum of sectional curvatures is less than $(\pi/(2r))^2$), then the Fréchet mean Σ is unique (Le, 1995). The support to ensure uniqueness can be very large. For example, for Euclidean spaces (with sectional curvature zero), or for non-Euclidean spaces with negative sectional curvature, the Fréchet mean is always unique.

If we have a sample $S_1, \ldots, S_N$ of i.i.d. observations available then the sample Fréchet mean is calculated by finding

$$\hat{\Sigma} = \arg\inf_{\Sigma} \sum_{i=1}^{N} d(S_i, \Sigma)^2.$$

Uniqueness of the sample Fréchet mean can also be determined from the result of Le (1995).

3.2 Distances between covariance matrices

We now consider specific choices of distances in order to provide estimates of a mean from the sample of N covariance matrices. To ensure the positive definiteness of the covariance matrices, a reparameterization can be used such as $S_i = Q_i Q_i^T$, where $Q_i \in R^{3\times 3}$. For example, $Q_i = chol(S_i)$ is the *Cholesky decomposition*, where Q_i is lower triangular with positive diagonal elements. Note that Q_i and any rotation and reflection of it $Q_i R_i$ ($R_i \in O(3)$) can result in the same S_i, i.e. $S_i = Q_i Q_i^T = Q_i R_i (Q_i R_i)^T, i = 1, ..., N$.

In applications there are several choices of distances between covariance matrices that one could consider, for example see Table 1.

Name	Notation	Form	Estimator
Euclidean	$d_E(S_1, S_2)$	$\|S_1 - S_2\|$	$\hat{\Sigma}_E$
Log-Euclidean	$d_L(S_1, S_2)$	$\|\log(S_1) - \log(S_2)\|$	$\hat{\Sigma}_L$
Riemannian	$d_R(S_1, S_2)$	$\|\log(S_1^{-1/2} S_2 S_1^{-1/2})\|$	$\hat{\Sigma}_R$
Cholesky	$d_C(S_1, S_2)$	$\|\text{chol}(S_1) - \text{chol}(S_2)\|$	$\hat{\Sigma}_C$
Root Euclidean	$d_H(S_1, S_2)$	$\|S_1^{1/2} - S_2^{1/2}\|$	$\hat{\Sigma}_H$
Procrustes size-and-shape	$d_S(S_1, S_2)$	$\inf_{R\in O(k)} \|\text{chol}(S_1) - \text{chol}(S_2)R\|$	$\hat{\Sigma}_S$
Full Procrustes shape	$d_F(S_1, S_2)$	$\inf_{R\in O(k),\beta\in\mathbf{R}} \left\|\frac{\text{chol}(S_1)}{\|\text{chol}(S_1)\|} - \beta\text{chol}(S_2)R\right\|$	$\hat{\Sigma}_F$
Power Euclidean	$d_A(S_1, S_2)$	$\frac{1}{\alpha}\|S_1^\alpha - S_2^\alpha\|$	$\hat{\Sigma}_A$

Table 1: *Some distances between covariance matrices and notation for the corresponding Fréchet mean estimators.*

Estimators $\hat{\Sigma}_E, \hat{\Sigma}_C, \hat{\Sigma}_H, \hat{\Sigma}_L, \hat{\Sigma}_A$ given in Table 1 are straightforward to compute using arithmetic averages. Note that d_S is obtained by optimal rotation/reflection of $chol(S_2)$ onto $chol(S_1)$ using ordinary Procrustes analysis. The Procrustes based estimators $\hat{\Sigma}_S, \hat{\Sigma}_F$ involve the use of the Generalized Procrustes Algorithm, which works well in practice (see Dryden et al., 2009). The Riemannian metric estimator $\hat{\Sigma}_R$ uses a gradient descent algorithm which is guaranteed to converge (e.g. see Pennec et al, 2006). In practice it is similar to the log-Euclidean estimator $\hat{\Sigma}_L$ (Arsigny et al., 2007).

We briefly summarize some of the properties of the distances. All these distances are invariant under simultaneous rotation and reflection of S_1 and S_2, i.e. the distances are unchanged by replacing both S_i by VS_iV^T, $V \in O(k), i = 1, 2$. Metrics $d_L(), d_R(), d_F()$ are invariant under simultaneous scaling of $S_i, i = 1, 2$, i.e. replacing both S_i by βS_i. Metric $d_R()$ is also affine invariant, i.e. the distances are unchanged by replacing both S_i by $AS_iA^T, i = 1, 2$ where A is a general $k \times k$ full rank matrix. Metrics $d_L(), d_R()$ have the property that $d(A, I_k) = d(A^{-1}, I_k)$, where I_k is the $k \times k$ identity matrix, and $d_L(), d_R(), d_F()$ are not valid for comparing rank deficient covariance matrices. Finally, there are problems with extrapolation with metric $d_E()$: extrapolate too far and the matrices are no longer positive semi-definite (Arsigny et al., 2007).

An alternative anisotropy measure to FA in (1) is to use the full Procrustes shape distance to isotropy where

$$PA = \sqrt{\frac{k}{k-1}} d_F(I_k, S) = \left\{ \frac{k}{k-1} \sum_{i=1}^{k} (\sqrt{\lambda_i} - \overline{\sqrt{\lambda}})^2 / \sum_{i=1}^{k} \lambda_i \right\}^{1/2},$$

where $\overline{\sqrt{\lambda}} = \frac{1}{k}\sum \sqrt{\lambda_i}$. We include the scale factor when defining the Procrustes Anisotropy (PA), and so $0 \le PA \le 1$, with $PA = 0$ indicating isotropy, and $PA \approx 1$ indicating a very strong principal axis.

Another anisotropy measure based on metrics d_L or d_R is the geodesic anisotropy

$$GA = \left\{ \sum_{i=1}^{k} (\log \lambda_i - \overline{\log \lambda})^2 \right\}^{1/2},$$

where $0 \leq GA < \infty$ (Arsigny et al., 2007), which has been used in diffusion tensor analysis in medical imaging with $k = 3$. Alternatively one could consider $\tanh(GA)$ (Batchelor et al., 2005) which is on the scale $[0, 1)$.

In some applications covariance matrices are close to being deficient in rank. For example when FA or PA are equal to 1 then the covariance matrix is of rank 1. The Procrustes metrics can easily deal with deficient rank matrices, which is a strong advantage of the approach.

4 Interpolation methods

4.1 Weighted Generalised Procrustes Analysis

Frequently in diffusion tensor imaging it is of interest to interpolate between sets of tensors. The weighted Fréchet sample mean of S_1, ..., S_N at N voxels with a certain distance function $d()$ is defined by:

$$\bar{S} = \arg\inf_{S} \sum_{i=1}^{N} w_i d(S_i, S)^2, \tag{2}$$

where the weights w_i are proportional to a function of the Euclidean distance between locations of the tensors (voxels), $0 \leq w_i \leq 1$ and $\sum_{i=1}^{N} w_i = 1$.

We choose d_S for the distance and then Weighted Generalized Procrustes analysis (WGPA) is proposed to obtain the weighted mean of S_1, ..., S_N. The objective of WGPA under rotation and reflection is to minimise a sum of weighted squared Euclidean norms S_{WGPA} which is given by

$$\begin{aligned} S_{WGPA}(S_1, ..., S_N) &= \inf_{R_1,...,R_N} \sum_{i=1}^{N} w_i \parallel Q_i R_i - \sum_{j=1}^{n} w_j Q_j R_j \parallel^2 \\ &= \inf_{R_1,...,R_N} \sum_{i=1}^{N} w_i \parallel (1 - w_i) Q_i R_i - \sum_{j \neq i} w_j Q_j R_j \parallel^2 \\ &= \inf_{R_1,...,R_N} \sum_{i=1}^{n} \frac{w_i}{(1 - w_i)^2} \parallel Q_i R_i - \frac{1}{(1 - w_i)} \sum_{j \neq i} w_j Q_j R_j \parallel^2 . \end{aligned} \tag{3}$$

Let $\hat{R}_i, i = 1, ..., N$ be the estimates of the rotation matrices. Then, the *WGPA mean tensor* is given by

$$\bar{S}_{WGPA} = \bar{Q}_{WGPA} \bar{Q}_{WGPA}^T, \tag{4}$$

where $\bar{Q}_{WGPA} = \sum_{i=1}^{N} w_i Q_i \hat{R}_i$. We give Algorithm 1 for estimating $\hat{R}_i, i = 1, ..., N$. Note that the algorithm is guaranteed to converge to a local minimum as the reduction in S_c at each iteration is at least zero.

Algorithm 1 Weighted Generalised Procrustes Method

Initial setting: $Q_i^P \leftarrow chol(D_i)$, $i = 1, ..., N$
S_{WGPA} from previous iteration: $S_p \leftarrow 0$
S_{WGPA} from current iteration: $S_c \leftarrow \sum_{i=1}^{N} w_i \parallel Q_i^P - \sum_{j=1}^{N} w_j Q_j^P \parallel^2$
while $|S_p - S_c| >$ tolerance **do**
  for $i = 1$ to N **do**
    $\bar{Q}_i = \frac{1}{1-w_i} \sum_{j \neq i} w_j Q_j^P$
    Calculate the rotation matrix R_i which minimises $\parallel \bar{Q}_i - Q_i^P R_i \parallel$ with partial ordinary Procrustes analysis
    $Q_i^P \leftarrow Q_i^P R_i$
  end for
  $S_p \leftarrow S_c$
  $S_c \leftarrow \sum_{i=1}^{N} w_i \parallel Q_i^P - \sum_{j=1}^{N} w_j Q_j^P \parallel^2$
end while
$\bar{Q}_{WGPA} \leftarrow \sum_{i=1}^{N} w_i Q_i^P$
return $\bar{Q}_{WGPA}$

4.2 Regularization

In medical image analysis a noisy tensor field may be available and so we wish to carry out regularization. For example, consider a grid of tensors $S_1, \ldots, S_n$ at voxels $x_1, \ldots, x_n$ and we wish to predict the tensor at a new site x. We could use the weighted penalized predictor obtained by minimizing, with respect to Σ,

$$\hat{\Sigma}_{\beta,\omega}(\lambda) = \sum_{i=1}^{n} w_i \text{dist}(S_i, \Sigma)^\beta + \lambda \text{dist}(\Sigma, \mu)^\omega$$

where the weights $w_i \geq 0, \sum w_i = 1$ are functions of the distance to the new site, $\lambda > 0$ is a regularization parameter, and μ is a reference matrix, such as the identity matrix, zero matrix or an overall average. For example we could use $w_i \propto \exp\{-\gamma \|x - x_i\|^2\}, i = 1, \ldots, n$.

Consider now smoothing across an image at the voxels $x_1, \ldots, x_n$, and so we need to minimize, with respect to $\Sigma_j, j = 1, \ldots, n$,

$$\sum_{j=1}^{n} \sum_{i=1}^{n} w_{ij} \text{dist}(S_i, \Sigma_j)^\beta + \lambda \sum_{j=1}^{n} \text{dist}(\Sigma_j, \mu)^\omega,$$

and w_{ij} is the weight as a function of the distance between sites i and j. Note $(\beta, \omega) = (2, 0)$ gives the weighted Fréchet mean, if $(\beta, \omega) = (\beta, 0)$ we have a type of M-estimator (Kent, 1992; Dryden and Mardia, 1998, p298), if $(\beta, \omega) = (1, 0)$ we have the geometric median (Fletcher and Joshi, 2009), if $(\beta, \omega) = (2, 2)$ non-Euclidean type of ridge-regression, and if $(\beta, \omega) = (2, 1)$ a non-Euclidean type of LASSO (see Tibshirani, 1996). Note that for the power metric (and Euclidean and square root) the space is Euclidean, and so using this procedure is relatively straightforward in this case.

5 Applications

5.1 Anisotropy of diffusion tensors

We consider anisotropy of estimated diffusion tensors in the brain obtained from diffusion weighted images (see Dryden et al., 2009). In Figure 1 we see a coronal view of the brain, and the corpus

callosum and cingulum can be seen.

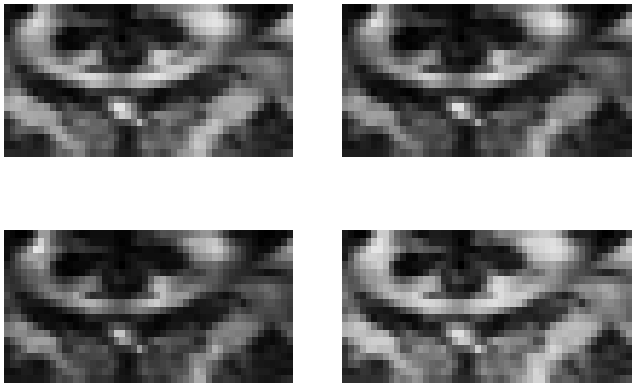

Figure 1: *The anisotropy measures (top left) FA, (top right) PA, (bottom left) GA and (bottom right)* $tanh(GA)$

At first sight all three anisotropy measures appear broadly similar. However, the PA image offers more contrast than the FA image in the highly anisotropic region - the corpus callosum. Also, the GA image has rather fewer brighter areas than PA or FA. The plot of $tanh(GA)$ is most different from the others, with much fewer dark areas. Due to the improved contrast we believe PA is slightly preferable in this example.

Figure 2: *Principal geodesic analysis for covariance matrices. The true geodesic path is given in the penultimate row (black). We then add noise in the three initial rows (red). Then we estimate the mean and find the first principal component (yellow), displayed in the bottom row.*

5.2 Principal geodesics of covariance matrices

We consider now an example estimating the principal geodesics of the covariance matrices $S_1, \ldots, S_n$ using the Procrustes size-and-shape metric d_S (see Dryden et al., 2009). Huckeman et al. (2009) discuss geodesic principal components analysis in Riemannian manifolds in depth. We consider an approximate procedure where the principal geodesics are estimated by principal components analysis of the tangent space co-ordinates. In Figure 2, we consider a true geodesic path (black) and evaluate 11 equally spaced covariance matrices along this path. We then add i.i.d. Gaussian noise in the tangent space for three separate realisations of noisy paths (in red). The overall mean $\hat{\Sigma}_S$ is computed

based on all the data ($n = 33$), and then the Procrustes size-and-shape tangent space co-ordinates are obtained based on the Cholesky decompositions of the covariance matrices. The first principal component loadings are computed and projected back to give an estimated minimal geodesic in the covariance matrix space. We plot this path in yellow by displaying 11 covariance matrices along the path. It can be seen that the estimated principal geodesic is very similar to the true geodesic path here. Other extensions include curve fitting through paths of covariance matrices using polynomials and geodesics (e.g. see Evans et al., 2009, for some examples of shape curves).

5.3 Interpolation

A tensor field from a healthy human brain has been smoothed and interpolated (with 2 interpolations between each pair of original voxels). The Fractional Anisotropy (FA) maps from the processed tensors are shown in Figure 3. Obviously, the FA map from the processed tensor data is much smoother than the one without processing. The feature that the cingulum is distinct from the corpus callosum is clearer in the anisotropy map from the processed data than those without processing in Figure 3.

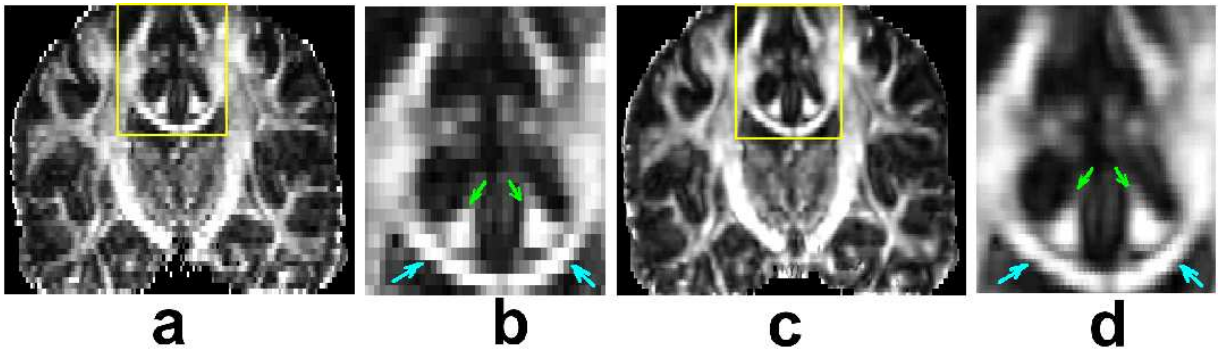

Figure 3: *Smoothing and interpolation of the diffusion tensor data from human brain. a: FA map from Bayesian tensor field. c: FA map from processed tensor field. b and d: Zoomed inset regions. Green arrows: the cingulum. Light blue arrows: the corpus callosum.*

5.4 Tractography

As a final application we give some initial results of fibre tractographies of the brain stem in a healthy human in Figure 4. It is of great interest to study the white matter fibre tracts in the brain in order to explore connectivity between different parts, both in healthy and patient brains. From different seed points in the brain stem, white matter fibres are tracked by following interpolated paths of principal directions from diffusion tensors. Tractography from the WGPA processed tensor field is different from the other methods, and work is currently underway to assess whether WGPA is preferable.

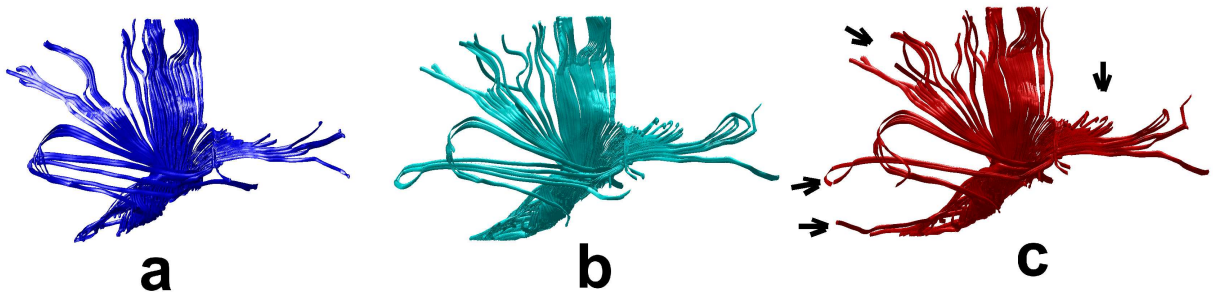

Figure 4: *Fibre tractograhpies using the Bayesian estimates (a), Euclidean smoothing (b) and WGPA smoothing (c). Black arrows point out some obvious differences of the WGPA tracts compared with other methods.*

6 Conclusions

Methodology for estimation and inference in the space of covariance matrices has application in many areas, including diffusion tensor imaging, structural tensor analysis in computer vision, and modelling longitudinal data with Bayesian and random effect models. There are many choices of metric available, each with its advantages. The particular choice of what is best will depend on the particular application. The use of the Procrustes size-and-shape metric d_S is particularly appropriate when the covariance matrices are close to being deficient in rank.

REFERENCES (RÉFERENCES)

Alexander, D. C. (2005). Multiple-fiber reconstruction algorithms for diffusion MRI. *Ann NY Acad Sci*, 1064:113–133.

Arsigny, V., Fillard, P., Pennec, A. and Ayache, N. (2007). Geometric Means in a Novel Vector Space Structure on Symmetric Positive-Definite Matrices, *SIAM Journal on Matrix Analysis and Applications*, **29**, 328–347.

Basser, P. J., Mattiello, J., and Le Bihan, D. (1994). Estimation of the effective self-diffusion tensor from the NMR spin echo. *J Magn Reson B.*, **103**, 247–254.

Batchelor P.G, Moakher M, Atkinson D, Calamante F, Connelly A. (2005). A rigorous framework for diffusion tensor calculus. *Magn. Reson. Med.*, **53**, 221–225.

Dryden, I. L., Koloydenko, A., and Zhou, D. (2009). Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *Annals of Applied Statistics*. To appear.

Dryden, I. L. and Mardia, K. V. (1998). *Statistical Shape Analysis*. Wiley, Chichester.

Evans, K., Dryden, I. L., and Le, H. (2009). Shape curves and geodesic modelling. Technical report, Division of Statistics, University of Nottingham. Submitted for publication.

Fillard, P., Arsigny, V., Pennec, X., and Ayache, N. (2007). Clinical DT-MRI estimation, smoothing and fiber tracking with log-Euclidean metrics. *IEEE Transactions on Medical Imaging*, **26**, 1472–1482.

Fletcher, P.T., Venkatasubramanian, S. and Joshi, S. (2009). The geometric median on Riemannian manifolds with application to robust atlas estimation. 45, S143-S152.

Fréchet, M. (1948). Les éléments aléatoires de nature quelconque dans un espace distancié. *Ann. Inst. H. Poincaré*, 10:215–310.

Huckemann, S., Hotz, T. and Munk, A. (2009). Intrinsic Shape Analysis: Geodesic Principal Component Analysis for Riemannian Manifolds Modulo Lie Group Actions. Discussion paper, Statistica Sinica, to appear.

Karcher, H. (1977). Riemannian center of mass and mollifier smoothing. *Comm. Pure Appl. Math.*, 30(5):509–541.

Kent, J. T. (1992). New directions in shape analysis. In Mardia, K. V., editor, *The Art of Statistical Science*, pages 115–127. Wiley, Chichester.

Le, H.-L. (1995). Mean size-and-shapes and mean shapes: a geometric point of view. *Advances in Applied Probability*, 27:44–55.

Pennec, X., Fillard, P. and Ayache, N. (2006). A Riemannian Framework for Tensor Computing, *Int. J. Comput. Vision*, **66**, 41–66.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. Royal. Statist. Soc B.*, Vol. 58, 26-7-288).

Zhou, D., Dryden, I. L., Koloydenko, A., and Bai, L. (2008). A Bayesian method with reparameterisation for diffusion tensor imaging. In Reinhardt, J. M. and Pluim, J. P. W., editors, *Proceedings, SPIE conference. Medical Imaging 2008: Image Processing*, page 69142J.